M7988 Modely ztrát v neživotním pojištění

Parametrický model a úlohy matematické statistiky

Model: $X_1, \dots, X_n$ je náhodný výběr z rozdělení s distribuční funkcí $F(x, \boldsymbol{\theta})$. Tuto distribuční funkci známe až na neznámý parametr $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$.

Úlohy matematické statistiky:

- Bodový odhad parametru $\boldsymbol{\theta}$.
- Intervalový odhad parametru $\boldsymbol{\theta}$.
- Testy hypotéz o parametru $\boldsymbol{\theta}$.

Někdy nás místo samotného parametru $\boldsymbol{\theta}$ zajímá nějaká jeho funkce, tzv. *parametrická funkce* $\gamma(\boldsymbol{\theta})$, kde $\gamma : \boldsymbol{\Theta} \to \mathbb{R}$ je reálná funkce.

Bodové odhady

Řekneme, že $T : \mathbb{R}^n \to \mathbb{R}$ je bodový odhad parametru $\theta \in \Theta \subset \mathbb{R}$, jestliže T je měřitelnou funkcí náhodného výběru $X_1, \dots, X_n$. Tedy $T = T(X_1, \dots, X_n)$ je náhodná veličina.

Vlastnosti bodových odhadů:

- T je nestranný odhad parametru θ, jestliže $\mathbb{E}T = \theta$ pro všechna $\theta \in \Theta$.
- T je konzistentní odhad parametru θ, jestliže $T = T(X_1, \dots, X_n) \to \theta$ v pravděpodobnosti pro $n \to \infty$ pro všechna $\theta \in \Theta$.

Který odhad je nejlepší?

- Nechť T_1, T_2 jsou dva nestranné odhady parametru θ. Řekneme, že T_1 je více eficientní (*lepší*) než T_2, jestliže $\mathrm{D}\,T_1 \leq \mathrm{D}\,T_2$ pro všechna $\theta \in \Theta$.
- Nechť T je nestranný odhad parametru θ. Řekneme, že T je nejlepší nestranný odhad parametru θ, jestliže $\mathrm{D}\,T \leq \mathrm{D}\,T^*$ pro všechna $\theta \in \Theta$ a pro všechny nestranné odhady T^*.
- Nechť T je odhad parametru θ. Střední čtvercovou (kvadratickou) chybu odhadu definujeme jako $\mathrm{MSE}(T) = \mathbb{E}(T - \theta)^2$.
- Je-li T je nestranný odhad parametru θ, pak $\mathrm{MSE}(T) = \mathrm{D}\,T$.
- Nechť T_1, T_2 jsou dva odhady parametru θ. Řekneme, že T_1 je více eficientní (*lepší*) než T_2, jestliže $\mathrm{MSE}(T_1) \leq \mathrm{MSE}(T_2)$ pro všechna $\theta \in \Theta$.
- (Stejnoměrně) nejlepší odhad parametru θ neexistuje.

Metoda momentů

- Dále předpokládejme, že neznámý parametr $\boldsymbol{\theta}$ je p-rozměrný ($\boldsymbol{\Theta} \subset \mathbb{R}^p$).
- Nechť existují obecné momenty $\mu'_k = \mu'_k(\boldsymbol{\theta}) = \mathbb{E}X_1^k$ pro $k = 1, \ldots, p$.
- Označme jejich výběrové protějšky $M'_k = \frac{1}{n}\sum_{i=1}^n X_i^k$ pro $k = 1, 2, \ldots$.
- Řekneme, že $\widetilde{\boldsymbol{\theta}}$ je odhad parametru $\boldsymbol{\theta}$ metodou momentů, jestliže $\mu'_k(\widetilde{\boldsymbol{\theta}}) = M'_k$ pro $k = 1, \ldots, p$.
- Je-li řešení předchozí soustavy nejednoznačné (rovnice jsou lineárně závislé), přidáme další rovnici pro $k = p + 1$, pokud ovšem existuje příslušný moment.

Metoda maximální věrohodnosti

- Označme sdruženou hustotu náhodného vektoru $(X_1, \dots, X_n)'$ jako

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta}).$$

- $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}, x_1, \dots, x_n)$ se nazývá věrohodnostní funkce.
- $\widehat{\boldsymbol{\theta}}$ se nazývá maximálně věrohodným odhadem parametru $\boldsymbol{\theta}$, jestliže $L(\widehat{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta}),\ \forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$.
- $\widehat{\boldsymbol{\theta}} = \arg\max\{L(\boldsymbol{\theta}); \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.
- $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i, \boldsymbol{\theta})$ se nazývá logaritmická věrohodnostní funkce.
- $\widehat{\boldsymbol{\theta}} = \arg\max\{l(\boldsymbol{\theta}); \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.

Vlastnosti maximálně věrohodných odhadů

- Za mírných předpokladů (podmínky regularity) jsou maximálně věrohodné odhady asymptoticky nestranné, konzistentní a mají asymptoticky normální rozdělení.
- $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ má asymptoticky rozdělení $\mathcal{N}_p(\mathbf{0}, \mathbf{J}^{-1}(\boldsymbol{\theta}))$.
- $$\mathbf{J}(\boldsymbol{\theta}) = \left(\mathbb{E} \frac{\partial \log f(X_1, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log f(X_1, \boldsymbol{\theta})}{\partial \theta_j} \right)_{i,j=1}^{p}$$

  je Fisherova informační matice o parametru $\boldsymbol{\theta}$ příslušná X_1.
- $$\mathbf{J}(\boldsymbol{\theta}) = -\mathbb{E} \left(\frac{\partial^2 \log f(X_1, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right)_{i,j=1}^{p}.$$
- $\widehat{\boldsymbol{\theta}} \approx \mathcal{N}_p(\boldsymbol{\theta}, \frac{1}{n}\mathbf{J}^{-1}(\boldsymbol{\theta}))$.

Metoda maximální věrohodnosti pro intervalová data

- Obor hodnot náhodného výběru je rozdělen na intervaly $(c_0, c_1], (c_1, c_2], \dots, (c_{k-1}, c_k]$.
- Označme n_j počet pozorování, které leží v intervalu $(c_{j-1}, c_j]$.
- Pravděpodobnost, že dané pozorování X_i leží v intervalu $(c_{j-1}, c_j]$ je

$$P(X_i \in (c_{j-1}, c_j]) = F(c_j, \boldsymbol{\theta}) - F(c_{j-1}, \boldsymbol{\theta}).$$

- $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} P(X_i \in (c_{j-1}, c_j]) = [F(c_1, \boldsymbol{\theta}) - F(c_0, \boldsymbol{\theta})]^{n_1} \cdot [F(c_2, \boldsymbol{\theta}) - F(c_1, \boldsymbol{\theta})]^{n_2} \cdot \dots \cdot [F(c_{k-1}, \boldsymbol{\theta}) - F(c_k, \boldsymbol{\theta})]^{n_k} = \prod_{j=1}^{k} [F(c_j, \boldsymbol{\theta}) - F(c_{j-1}, \boldsymbol{\theta})]^{n_j}$.
- $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{j=1}^{k} n_j \log[F(c_j, \boldsymbol{\theta}) - F(c_{j-1}, \boldsymbol{\theta})]$ se nazývá logaritmická věrohodnostní funkce.
- $\widehat{\boldsymbol{\theta}} = \arg\max\{l(\boldsymbol{\theta}); \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.

Maximálně věrohodné odhady pro parametrickou funkci γ

- Nechť $\gamma : \mathbf{\Theta} \to \mathbf{\Theta}^*$ je prostá parametrická funkce.
- Funkci $\widetilde{L}(\boldsymbol{\theta}^*) = \prod_{i=1}^n f(x_i, \gamma^{-1}(\boldsymbol{\theta}^*))$ pro $\boldsymbol{\theta}^* \in \mathbf{\Theta}^*$ nazveme věrohodnostní funkcí indukovanou parametrickou funkcí γ.
- $\widehat{\boldsymbol{\theta}}^*$ je maximálně věrohodný odhad parametrické funkce $\gamma(\boldsymbol{\theta}) = \boldsymbol{\theta}^*$, jestliže $\widetilde{L}(\widehat{\boldsymbol{\theta}}^*) \geq \widetilde{L}(\boldsymbol{\theta}^*)$ pro všechna $\boldsymbol{\theta}^* \in \mathbf{\Theta}^*$.
- Zehnaova věta (princip invariance MLE): Je-li $\widehat{\boldsymbol{\theta}}$ maximálně věrohodný odhad parametru $\boldsymbol{\theta}$, pak $\gamma(\widehat{\boldsymbol{\theta}})$ je maximálně věrohodný odhad parametrické funkce $\gamma(\boldsymbol{\theta})$.

Delta metoda

Theorem

Nechť $\{\mathbf{X}_n\}_{n=1}^{\infty}$ *je posloupnost p-rozměrných náhodných vektorů takových, že* $\sqrt{n}(\mathbf{X}_n - \boldsymbol{\theta})$ *má asymptoticky normální rozdělení* $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. *Dále buď* $\gamma : \mathbb{R}^p \to \mathbb{R}$ *měřitelná funkce, která má totální diferenciál v bodě* $\boldsymbol{\theta}$. *Pak platí:* $\sqrt{n}(\gamma(\mathbf{X}_n) - \gamma(\boldsymbol{\theta}))$ *má asymptoticky normální rozdělení* $\mathcal{N}(0, \sigma^2)$, *kde* $\sigma^2 = \nabla\gamma'(\boldsymbol{\theta})\boldsymbol{\Sigma}\nabla\gamma(\boldsymbol{\theta})$, *kde*

$$\nabla\gamma(\boldsymbol{\theta}) = \left(\frac{\partial\gamma(\boldsymbol{\theta})}{\partial\theta_1}, \dots, \frac{\partial\gamma(\boldsymbol{\theta})}{\partial\theta_p}\right)'$$

je gradient funkce γ *v bodě* $\boldsymbol{\theta}$.

Aplikace na MLE

- Víme, že za mírných předpokladů (podmínky regularity) platí: $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ má asymptoticky rozdělení $\mathcal{N}_p(\mathbf{0}, \mathbf{J}^{-1}(\boldsymbol{\theta}))$.
- Aplikací delta metody dostaneme, že za mírných předpokladů platí: $\sqrt{n}(\gamma(\widehat{\boldsymbol{\theta}}) - \gamma(\boldsymbol{\theta}))$ má asymptoticky rozdělení $\mathcal{N}(0, \nabla\gamma'(\boldsymbol{\theta})\mathbf{J}^{-1}(\boldsymbol{\theta})\nabla\gamma(\boldsymbol{\theta}))$.
- $\gamma(\widehat{\boldsymbol{\theta}}) \approx \mathcal{N}(\gamma(\boldsymbol{\theta}), \frac{1}{n}\nabla\gamma'(\boldsymbol{\theta})\mathbf{J}^{-1}(\boldsymbol{\theta})\nabla\gamma(\boldsymbol{\theta}))$.
- Je-li θ jednorozměrný parametr, pak $\gamma(\widehat{\theta}) \approx \mathcal{N}\left(\gamma(\theta), \frac{[\gamma'(\theta)]^2}{nJ(\theta)}\right)$.

Metoda minimálního χ^2

1. Rozděl obor hodnot náhodné veličiny X_i na k po dvou disjunktních intervalů $B_1, \dots, B_k$.
2. Označ $Y_j = \sum_{i=1}^n \mathbb{I}\{X_i \in B_j\}$ pro $j = 1, \dots, k$ počet pozorování, které padnou do intervalu B_j.
3. Spočítej teoretickou pravděpodobnost, že náhodná veličina X_i nabude hodnoty z intervalu B_j = pravděpodobnost, že dané pozorování padne do intrvalu B_j: $p_j(\boldsymbol{\theta}) = P(X_i \in B_j) = \int_{B_j} f(x, \boldsymbol{\theta})$.
4. Porovnej očekávaný a skutečný (pozorovaný) počet pozorování v jednotlivých intervalech B_j pomocí Pearsonovy χ^2 statistiky testu dobré shody:
$$\chi^2(\boldsymbol{\theta}) = \sum_{j=1}^k \left(\frac{Y_j - np_j(\boldsymbol{\theta})}{\sqrt{np_j(\boldsymbol{\theta})}} \right)^2 .$$
5. Odhad parametru $\boldsymbol{\theta}$ metodou minimálního χ^2 minimalizuje $\chi^2(\boldsymbol{\theta})$ přes všechny hodnoty $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, tj. $\widetilde{\boldsymbol{\theta}} = \arg\min\{\chi^2(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.

Metoda minimálního χ^2 - poznámky

- $$\chi^2(\boldsymbol{\theta}) = \sum_{j=1}^{k} \frac{Y_j^2}{np_j(\boldsymbol{\theta})} - n.$$
- Intervaly B_j by se měly volit stejně pravděpodobné, tj. $p_j(\widetilde{\boldsymbol{\theta}}) = \frac{1}{k}$ pro $j = 1, \ldots, k$.
- Volba počtu tříd k - heuristická pravidla, např. $k \doteq 15 \left(\frac{n}{100}\right)^{2/5}$, nebo $k \doteq 2n^{2/5}$.

Bayesovské odhady (Bayesovská statistika)

- Kombinuje informaci obsaženou v datech (parametrický model) s apriorní informací o neznámém parametru θ (zkušenosti, domněnky, dřívější pozorování).
- Závěry (odhady) vyvozuje až z aposteriorního rozdělení.
- Idea: Naše informace o hodnotě neznámého parametru může být vyjádřena pomocí pravděpodobnostního rozdělení, tj. neznámý parametr θ považujeme za náhodnou veličinu.

Matematický model

- $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s hustotou $f(x, \boldsymbol{\theta})$, kde $\theta \in \Theta$.
- θ je nyní náhodná veličina s hustotou $q(\theta)$.
- Označme podmíněnou hustotu náhodného vektoru $(X_1, \ldots, X_n)'$ při dané hodnotě parametru θ jako $r(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i, \theta)$, kde $\mathbf{x} = (x_1, \ldots, x_n)'$.

Theorem (Bayesova věta)

Pro podmíněnou hustotu náhodné veličiny θ při daných hodnotách $\mathbf{X} = \mathbf{x}$ platí:

$$\pi(\theta|\mathbf{x}) = \begin{cases} \frac{r(\mathbf{x}|\theta)q(\theta)}{\int_\Theta r(\mathbf{x}|\theta)q(\theta)d\theta}, & \textit{pokud } \int_\Theta r(\mathbf{x}|\theta)q(\theta)d\theta \neq 0, \\ 0, & \textit{jinak.} \end{cases}$$

Poznámky

- $q(\theta)$ se nazývá apriorní hustota - vyjadřuje informaci o parametru θ ještě před realizací náhodného výběru $\mathbf{X}$.
- $\pi(\theta|\mathbf{x})$ se nazývá aposteriorní hustota - vyjadřuje informaci o parametru θ až po realizaci náhodného výběru $\mathbf{X}$.
- Při Bayesovském přístupu používáme kromě dat (realizace náhodného výběru) ještě informaci o parametru θ nezávisle na našich datech.
- Tato informace může mít objektivní i subjektivní charakter.

Volby apriorního rozdělení

- Pokud máme informace (výsledky) z minulosti
  - přesná znalost rozdělení
  - jádrové odhady hustoty
  - parametrický model
- Pokud nemáme informace (výsledky) z minulosti
  - neinformativní (rovnoměrné) rozdělení
  - Jeffreysovo apriorní rozdělení $q(\boldsymbol{\theta}) \sim \sqrt{|\mathbf{J}(\boldsymbol{\theta})|}$
  - konjugované apriorní rozdělení

Bodové odhady

- Definujme ztrátovou funkci $L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ - ztráta, kterou utrpíme, když odhadneme parametr $\boldsymbol{\theta}$ pomocí odhadu $\widehat{\boldsymbol{\theta}}$.
- Dále definujme bayesovské riziko (průměrná aposteriorní ztráta):

$$r(\boldsymbol{\theta}) = \int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) \pi(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta}.$$

- Hledáme odhad, který minimalizuje bayesovské riziko.
- Pro kvadratickou ztrátovou funkci $L(\theta, \widehat{\theta}) = (\theta - \widehat{\theta})^2$ je bayesovským odhadem aposteriorní střední hodnota, tj. $\widehat{\theta} = \mathbb{E}(\theta | X_1, \ldots, X_n)$.
- Pro absolutní ztrátovou funkci $L(\theta, \widehat{\theta}) = |\theta - \widehat{\theta}|$ je bayesovským odhadem aposteriorní medián, tj. $\widehat{\theta} = \mathrm{med}(\theta | X_1, \ldots, X_n)$.
- Pro 0-1 ztrátovou funkci $L(\theta, \widehat{\theta}) = \mathbb{I}\{\theta \neq \widehat{\theta}\}$ je bayesovským odhadem aposteriorní modus, tj. $\widehat{\theta} = \arg\max \pi(\theta | X_1, \ldots, X_n)$.

Intervalové odhady

Definition

$100(1-\alpha)\%$ věrohodnostní interval pro parametr θ je takový interval $[a,b] = [a(X_1,\dots,X_n), b(X_1,\dots,X_n)]$, pro který $P(a \leq \theta \leq b|\mathbf{X}) = 1-\alpha$.

- Equal tail - je takový interval $[a,b]$, kde a je $\alpha/2$-kvantil aposteriorního rozdělení $\theta|\mathbf{X}$ a b je $1-\alpha/2$-kvantil aposteriorního rozdělení $\theta|\mathbf{X}$.
- HPD (interval o největší aposteriorní hustotě) je takový interval $[a,b]$ takový, že $\pi(\theta|\mathbf{x}) \geq c$ pro všechna $\theta \in [a,b]$ a $c > 0$ je nejmenší číslo takové, že $P(\pi(\theta|\mathbf{x}) \geq c) = 1-\alpha$.

Theorem

Je-li $\pi(\theta|\mathbf{x})$ spojitá a unimodální, pak HPD interval je nejkratší mezi všemi věrohodnostními intervaly.

Predikce budoucího pozorování

- Nechť se budoucí pozorování X_{n+1} řídí stejným modelem jako $X_1, \ldots, X_n$ - tedy má hustotu $f(x, \boldsymbol{\theta})$.
- Chceme predikovat(předpovídat) jeho budoucí hodnotu.
- S pomocí Bayesovy věty můžeme odvodit aposteriorní prediktivní hustotu

$$f(x_{n+1}|\mathbf{x}) = \int_{\boldsymbol{\Theta}} f(x_{n+1}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.$$

Model selection (výběr modelu)

1. Je náš model vhodný? Popisuje dobře naše data?
2. Máme-li více modelů, který z nich je nejlepší? Který máme použít?

Grafické metody pro posouzení vhodnosti modelu

- jsou založené na porovnání teoretického a empirického rozdělení
- porovnání teoretické a empirické distribuční funkce (hustoty) v jednom grafu
- Q-Q plot
- P-P plot

Kolmogorovův - Smirnovův test

- Nulová hypotéza H_0: $F = F^*$, kde F^* je známá distribuční funkce.
- Definujme empirickou distribuční funkci

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}.$$

- Testová statistika $D_n = \max_{x \in \mathbb{R}} \{|\widehat{F}_n(x) - F^*(x)|\}$.
- Za platnosti H_0 má $\sqrt{n} D_n$ asymptotické rozdělení stejné jako $\sup_{t \in [0,1]} |B(t)|$, kde $B(t)$ je Brownův most v $\mathcal{C}(0,1)$.
- Test se dá použít jen, když F^* je plně specifikována modelem. Pokud jsou k jejímu odhadu použita data, test nefunguje - je příliš konzervativní - nutná jeho modifikace, např. Lillieforsova.

Pearsonův χ^2 test dobré shody

- Nulová hypotéza H_0: $F = F^*$, kde F^* je známá distribuční funkce.
- Definujme si intervaly $(c_{j-1}, c_j]$, $j = 1, \dots, k$.
- Označme n_j počet pozorování, které padnou do intervalu $(c_{j-1}, c_j]$ pro $j = 1, \dots, k$.
- Určíme očekávaný počet pozorování, které by měly padnout do intervalu $(c_{j-1}, c_j]$:

$$e_j = np_j = nP(X_1 \in (c_{j-1}, c_j]) = n(F^*(c_j) - F^*(c_{j-1})).$$

- Testová statistika

$$\chi^2 = \sum_{j=1}^{k} \frac{(e_j - n_j)^2}{e_j}.$$

- χ^2 má za platnosti nulové hypotézy asymptoticky χ^2 rozdělení s $k-1$ stupni volnosti.

Modifikace Pearsonova χ^2 testu dobré shody

- Nulová hypotéza H_0: F je distribuční funkce nějakého rozdělení.
- Nejprve z dat odhadneme neznámý p-rozměrný parametr $\boldsymbol{\theta}$ a postupujeme stejně jako v předchozím.
- Upravená testová statistika má za platnosti nulové hypotézy asymptoticky χ^2 rozdělení s $k-1-p$ stupni volnosti.

Volba tříd:

- $p_j = \frac{1}{k}$ pro $j = 1, \dots, k$.
- Heuristická pravidla pro počet tříd – $k \doteq 2n^{2/5}$, nebo $k \doteq 15(n/100)^{2/5}$.

Výběr modelu z několika kandidátů

Princip Occamovy břitvy - vybíráme co nejjednodušší vhodný model.

1. Judgement-based přístup - založený na subjektivním úsudku analytika
   - rozhodnutí založené na různých grafech či tabulkách (tail vs. mod fit)
   - rozhodnutí založené na předchozí zkušenosti (Paretovo rozdělení pro výši příjmů, Benfordovo pro četnost prvních číslic)
   - model je plně určen situací, kterou má popisovat (házení mincí - alternativní rozdělení)
2. Score-based přístup - založený na číselných charakteristikách
   - nejnižší hodnota statistiky nějakého statistického testu
   - nejvyšší p-hodnota nějakého statistického testu
   - nejvyšší hodnota věrohodnosti
   - nejvyšší hodnota nějaké penalizované funkce, např. AIC, BIC